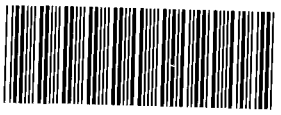
AD106

1 of 13



000064259

Date: March 10, 1995

To: File

From: NAH Neil Holsteen, OU 6 Project Manager

Subject: DATABASE DECISIONS FOR ECOLOGICAL RISK ASSESSMENT

This letter to file will document the decision process concerning database clean up for the ecological risk assessment.

The starting assumption concerning the analytical database for the ecological database was that the cleaned up data from each of the OUs would be consolidated into one database. For this to be successful, each database would be cleaned up in a similar manner and maintained in a similar structure. Unfortunately, this was not the case. The databases provided by the subcontractors for each OU are slightly different.

To facilitate the completion of the watershed ERA's, the Project Manager's chose to re-extract the data for each OU within the watersheds, and clean them up in a consistent manner.

1/13

March 9, 1995
2501-95/17

Neil Holsteen, Project Manager
EG&G Rocky Flats, Inc.
Building 080, Interlocken
P.O. Box 464
Golden, CO 80402-0464

Dear Mr. Holsteen:

As you know, Stoller is currently evaluating alternative sources of data for use in the ecological risk assessments (ERAs) for the Woman Creek and Walnut Creek watersheds at RFETS. Three sources have been considered:

1) the combined data sets obtained from contractors performing the "nature and extent" and human health risk assessment components of the RCRA Facility Investigation/Remedial Investigation (RFI/RI) for each operable unit

2) the RFETS database maintained by PRC for the Environmental Protection Agency (EPA)

3) raw data obtained directly from the RFEDS database maintained by EG&G

The primary issues in determining the usability of the data sets involve the extent of "cleanup" the data set has undergone. Cleanup is necessary because the data sets obtained from laboratories include data from quality assurance/quality control (QA/QC) samples that are not generally used in estimating concentrations of chemicals in environmental media. In addition, the adjustments must be made for the way certain parameters (e.g., non-detect results and detection limits) are reported by the laboratories.

EPA and EG&G subcontractors have performed the cleanup on various data sets using different methods and assumptions. These differences must be identified and reconciled if we are to aggregate the data for use in a watershed-scale assessment. The major issues to be addressed in this process are the extent to which the data has been "cleaned up," the process used to perform the cleanup, the extent to which the cleanup has been documented, and the level of effort remaining to reconcile differences.

Stoller

Please do not hesitate to call me if you have any questions. Thank you.

Sincerely,

Mark C. Lewis, Ph.D.
Project Manager

Enclosure

cc:	E. Mast	EG&G
	F. Vertucci	EG&G
	A. Crockett	Stoller
	L. Ross	Stoller
	MCL-Chron	Stoller

Over the past two weeks, Stoller has reviewed the data sets obtained by other subcontractors and performed some initial cleanup. Based on this review and discussions among EG&G, EPA, PRC, and Stoller personnel, it appears that extracting a new data set from RFEDS and then performing uniform cleanup procedure will result in the most defensible data set and can be accomplished sooner than either of the other two options. As we discussed, Stoller will proceed with obtaining database records to be used in the Woman Creek and Walnut Creek ERAs. We intend to make our RFEDS request specific to the needs of the ERAs. The request will be made using the following criteria:

- We will request data only for those locations used in the RFI/RI nature and extent and human health risk assessments. We have developed lists of location codes based on the data files received from other subcontractors responsible for the main RFI/RI reports.

- We will request data for the following sample types: borehole (BH), surface soil (SS), groundwater (GW), sediment (SD), and surface water (SW). We presently have all of the biological tissue data and will not need to request a new extraction.

- We will request data collected after January 1, 1990.

- Our request will exclude data records for tentatively identified compounds (TICs) (i.e., result type or secondary result type = TIC) and surrogate results (secondary result type = SURR).

- We will request only records with a QC code = REAL.

Once we have the RFEDS extraction, we will evaluate the data set using in-house dBASE programs prepared by Karen Schneider that will identify the types of cleanup operations that will be required. This information will then be used to guide cleanup routines needed to prepare the data set for use. Enclosed you will find a summary of the cleanup procedures. We would like you to review and comment on the extraction criteria and cleanup routines we have proposed.

5

DATA MANIPULATION FOR RFI/RI REPORTS
Suggested Handling for the Sitewide Ecological Risk Assessment

I. RFEDS EXTRACTION

A. Run DATA_CHL.PRG to get an overview of the data. Modifications to other programs (notably CLEANGEN.PRG) may be necessary if there are new/unusual data.

DATA_CHK.PRG performs the following operations:

1. Assigns an analytical group based on test group code: M for metals; P for pesticides, herbicides, and PCBs; R for radionuclides; S for semivolatile organic compounds; V for volatile organic compounds; and W for water quality parameters.

2. Creates text files with the results of data checks.

   a. Unique sample types.
   b. Unique analyte names, by analytical group. Used to identify multiple spellings of the same analyte.
   c. Unique QC codes for each sample type.
   d. Unique test group codes for each analytical group.
   e. Unique units for each analytical group and test group code.
   f. Unique result types for each analytical group and test group code.
   g. Unique lab qualifiers for each analytical group and test group code.
   h. Unique validation codes for each analytical group and test group code.
   i. The total number of records and the total number of unique RFEDS sequence identification numbers to check for duplicated records.
   j. Count the number of, and identify the records that have a blank result field.
   k. Count the number of, and identify the non-rads records that have a blank detection limit field.

Results from DATA_CHK.PRG are used to make any necessary changes to CLEANGEN.PRG.

B. The generic data cleanup routine program CLEANGEN.PRG is run on data files to:

1. Assign total/dissolved, organic/inorganic based on test group code.

2. Identifies and flags unusable data. Criteria under which a record may be rejected include:

   a. Blank results
   b. Blank units
   c. Blank detection limit and non-rads analysis
   d. Unknown validation codes (C, P, S, B, N)
   e. Unknown units (DPM/SA, PCI/SA, ROM BA, UNKN)
   f. Unknown lab qualifiers (L, R, rad UE, organic E)
   g. Alpha characters in numeric field
   h. Tentatively identified compound (TIC) laboratory qualifiers (A, organic N)

3. Assign a usability category based on validation codes and laboratory qualifiers:

   a. VALID:
      validation codes: A,V
      laboratory qualifiers: blank,U
   b. ESTIMATE:
      validation codes: A,J,V,JA
      laboratory qualifiers: +, *, B, C, D, E(inorganic), F, G, H, I, J, N, S, UJ, UN, UW, UX, W, X, Y, Z
   c. REJECT:
      validation codes: B,C,N,P,R,S
      laboratory qualifiers: E(organic),L,R,UE(rads)
   d. BLNK/Y VAL:
      validation codes: Y,blank
      laboratory qualifiers: blank, +, *, B, C, D, E(inorganic), F, G, H, I, J, N, S, U, UJ, UN, UW, UX, W, X, Y, Z
   e. CHECK:
      validation codes: Z
      laboratory qualifiers: all except rejection qualifiers

4. Make all units for a given analyte consistent:

   a. Solids:
      metals or Water Quality Parameters MG/KG
      radionuclides PCI/G
      organic compounds UG/KG
   b. Liquids:
      radionuclides PCI/L
      all other analytical groups UG/L

c. Convert DPM/L to PCI/L for "historical" rads results.

5. Assign detect or non-detect flag based on following criteria:

a. Y (detect):
all rads
other analytical groups:
validation codes: blank, Y, J, A, V
laboratory qualifiers: all except U, and those that reject the record

b. N (non-detect):
All rejected records
Validation codes: all
laboratory qualifiers: U, JB

C. The program NEW_DLIM.PRG is run to modify records with detection limits reported as the contract required detection limit (CRDL) (especially metals) instead of the instrument detection limit (IDL). The CRDL is usually reported in the DETECTION LIMIT field of non-detect (U-qualified) records. In these cases, the IDL is reported in the RESULT field (Paul Gomez, EG&G Rocky Flats, personal communication).

NEW_DLIM.PRG modifies the entry in the detection limit field.

1. If the reported detection limit is the CRDL for that analyte, entry in the detection limit field is replaced with the IDC from the reported result field. Otherwise, the new detection limit field is filled with the reported, converted detection limit.

D. The program NON_TAL.PRG removes records for analytes not in the combined list of PCOCs (the target analyte list for this project) from all data sets. NON_TAL.PRG performs the following operations:

1. Marks non-target records in the final structure files

2. Creates files containing those records removed

3. Checks those non_target analytes that aren't labeled as TICs for detects.

a. Creates a list of detected analytes. Includes the following fields: location, s_no, t_g_c, analyte, result, d_limit, qual, valid.

7

II. CREATION OF FINAL WORKING DATA SETS

A. The program SPLIT.PRG is run to sort the original database:

1. Records labeled "CHECK" (validation code of Z) are removed. RFEDS assigns the Z validation code based on the following criteria:
   - samples analyzed at on-site labs
   - geophysical samples
   - sample numbers starting with NP or VW
   - all laboratory qa samples that are typically stored in a separate RFEDS database
   - records with a blank result field and information in the laboratory disposition field

2. Unusable records (category = REJECT; usability = REJECT) are removed.

3. Records labeled as "TIC", based on certain laboratory qualifiers, are removed.

4. Records labeled as category REAL or QC with a usability code of VALID, ESTIMATE, or BLNK/Y VAL are kept together to go through the next cleanup steps.

B. The file resulting from SPLIT.PRG (REAL/QC) is checked for extra and laboratory QA/QC records that, if there is an associated target record, could make duplicate records for a given analyte. The records not chosen are removed. DUPERTYP.PRG is run iteratively to remove duplicate records.

1. Sorts file on s_no+t_g_c+analyte+r_type.

2. If none of the records are validated:
   keep the TRG record
   remove the other(s)

3. If one record is validated and the other(s) are not (i.e. valid is blank or Y):
   keep the validated record
   remove the other(s)

C. If more than one record for a given sample is validated the records are checked by hand and the decision to keep or remove is made on a record by record basis. Those records removed from the data set using these steps are also stored in a separate file LAB_QA.DBF.

9

MEMORANDUM

cc: Neil Hdsteen, EG&G

To: Mark Lewis
From: Karen Schneider
Date: 3/7/95
Re: Data Handling for the Sitewide Ecological Risk Assessment

The following issues were discovered during the combining of data sets from various contractors and OUs. Once final structures were created and overviews of the data generated, it became clear that additional "cleanup" steps may be necessary to create an internally consistent database.

1. Some data sets have the result (error and detection limit) columns converted to numeric data type. Numeric data types require a fixed number of decimal places. If the number of decimal places was set to too few, the information in the field(s) is rounded off. This can be a problem with radionuclide data where there is information out in the sixth or seventh decimal place. For example, if RFEDS reported the result as 0.000029 pci/l but the field was set to 4 decimal places, the database result would become 0.0000 pci/l.

2. Unclear in some data sets, by name and contents, which field is the detection limit and which the error for radionuclide data.

3. Most data sets included records with one or more of the following fields blank: result, detection limit, units. I suggest removing all of these records.

   A cursory check on a few records with blank results indicated that RFEDS has results for those records. Blank detection limits are a problem when there are non-detects. Typically, when means are calculated, non-detect results are replaced with the detection limit or 1/2 the detection limit. It is impossible to tell the "magnitude" of the concentration without units.

4. Tentatively identified compounds are included. I suggest removing these records based on the result type or secondary result type of "TIC".

5. Records for field qc samples are included, e.g. RNS, DUP, etc. I suggest removing these records and dealing only with REAL samples.

6. There are records with undefined lab qualifiers and validation codes. I suggest removing these records.

7. Records having a Z validation code should be removed. RFEDS assigns a Z based on the following criteria: samples analyzed at on-site labs, geophysical samples, sample numbers starting with VW or NP, lab QA records typically stored in the RFEDS VAL_QA table, records with a blank result field and information in the lab disposition field.

9

8. There are duplicate records for a given sample number, test group code, and analyte combination based on different result types, e.g. TRG, DIL, REX, etc. A single record must be selected for each sample number, test group code and analyte combination. Extra records should be removed. This is an iterative task.

9. There are multiple names for the same analyte. Consistent analyte names must exist and are based on the parameter code (casno). The parameter code is missing from some of the data sets.

10. Units must be made consistent for the same analytical group across all data sets.

11

SITEWIDE ECOLOGICAL RISK ASSESSMENT
STEPS FOR DATA HANDLING

Date Done	Step	
2/23/95	1.	Print out the Stoller "final" dbf structure.
2/28/95	2.	The final structure is in s:\eras\dbf\final.dbf. Add the following fields to the final structure:

Name	Width	
OU	2	(1,2,4,5,6,7,11)
DRAINAGE	6	(WALNUT, WOMAN, BOTH)
SORCE_AREA	20	(

2/23/95 3. Each person responsible for an individual Operational Unit's Data.

OU	Person	
1	Paul Schock	x358
2	Martha Plank	x419
4	Dayna Rigor	x453
5	Katharine Misken	x445
6	Ricky Bell	x447
7	Maureen O'Shea-Stone	x466
11	Martha Plank	x419

4. Set up individual systems as outlined below:

- Create an OUx directory
- Create an ORIG subdirectory
- Create a FINAL subdirectory
- Copy all of the original contractor files into the ORIG subdirectory.
- Zip all of the original files

5. Print a list of the filenames that each contractor delivered. 1 copy goes to Wym Chromac, 1 copy stays in the project file.

3/6/95	OU1
2/27/95	OU2
3/7/95	OU4
2/27/95	OU5
	OU6
3/7/95	OU7
2/27/95	OU11

2/27/95 6. Print out the structure and a sample of field contents for each file.

2/27/95 7. On the structure printout, identify the "Stoller" name for each contractor field. Make a copy of these sheets for the project file.

2/28/95 8. Identify on a representative (one for each structure type) structure the max width for every field.

2/28/85 9. Verify that all fields are being appropriately named, compare with all other structures and OUs.

2/28/95 10. Modify the contractor file structures and change field names to match the hand modified structure.

3/1/95 11. Check the modified structure against the Stoller final field names. The field names must match exactly. If they don't, modify structure again and correct the contractor file structures to match.

12

3/1/95 12. Copy the Stoller final structure (s:\eras\dbf\final.dbf) to individual machines into the OUx and FINAL subdirectory. Do the following steps at the dot prompt:

```
SET DIRECTORY TO C:\OUx\FINAL
USE FINAL
COPY STRU TO file1 - x          (create a "final" structure for each of the contractor files)
```

3/1/95 13. Append each of the contractor files into the corresponding final file.

```
SET DIRECTORY TO C:\OUx\FINAL
USE file1
APPEND FROM C:\OUx\ORIG\file1
```

3/2/95 14. Check the final file to make sure that the contents of every "important" field was indeed filled. If not, there is probably a problem with the field name. Check the field names, modify the ORIG file structure, and repeat step 13.

3/2/95 15. Once all fields are being filled properly, repeat step 13 for each ORIG/FINAL file.

3/2/95 16. Zip the contractor files in the ORIG directory.

3/2/95 17. Zip the Stoller structure files in the FINAL directory.

ALL THE FOLLOWING STEPS ARE PERFORMED ON THE STOLLER FINAL STRUCTURE FILES!

3/2/95 18. Add information contained in the filename to the records using the appropriate program. Not all filenames have this information. Fields in those files will be filled at a later date (not overwriting what has already been entered).

Field Info	Program
Total or Dissolved	CONT_TD.PRG
Analytical Group (M,R,...)	CONT_GRP.PRG

3/2/95 19. Zip the Stoller structure files in the FINAL directory.

3/3/95 20. Create the final OUx database. Use the following commands at the dot prompt.

```
SET DIRECTORY TO C:\OUx
USE FINAL
COPY STRU TO OUx_FINL
USE OUx_FINL
APPEND FROM C:\OUx\FINAL\file1...
```

3/3/95 21. Add the OU number to the FINAL file. Get the CONT_OU.PRG program from s:\eras\programs\; modify the program to match each OU; run the program at the dot prompt:

```
SET DIRECTORY TO C:\OUx
DO CONT_OU
```

3/3/95 22. Add the DRAINAGE to the FINAL file. Get the CONT_DRN.PRG program from s:\eras\programs\; modify the program to match each OU; run the program at the dot prompt:

```
SET DIRECTORY TO C:\OUx
DO CONT_DRN
```


12

3/6/95 23. Get an overview of the data for each OU. Get the DATA_CHK.PRG program from s:\eras\programs\; modify the program to match each OU; run the program at the dot prompt:

SET DIRECTORY TO C:\OUx
DO DATA_CHK

24. Print out a 2 copies of the OUx_DATA.TXT file.

3/6/95 OU1
3/7/95 OU2
3/6/95 OU4
3/7/95 OU5
OU6
3/7/95 OU7
3/7/95 OU11

25. Meet to compare the DATA_CHK.PRG output files to come up with a list of consistent analyte names, units, etc.

26. Write and run programs to modify OU databases to make consistent.

27. Add the source area to the OUx file. Get the CONTSRCE.PRG program from s:\eras\programs\; modify the program to match each OU; run the program at the dot prompt.

SET DIRECTORY TO C:\OUx
DO CONTSRCE

28. Run SPLIT.PRG to split the OU databases into the final media databases on the network, e.g. SERA_GW.DBF, SERA_SS.DBF, etc.

29. Copy the media databases to the appropriate media subdirectory in the s:\eras directory.